**PROPOSAL PENELITIAN**

**ANALISA BERMACAM OPTIMIZER PADA *CONVOLUTIONAL NEURAL NETWORK* UNTUK DETEKSI PENYAKIT JANTUNG**



**Oleh:**

**Fikri Ainun Najib**

**200411100153**

**PROGRAM STUDI INFORMATIKA**

**JURUSAN TEKNIK INFORMATIKA**

**FAKULTAS TEKNIK**

**UNIVERSITAS TRUNOJOYO MADURA**

**2022**

**BAB I**

**PENDAHULUAN**

**1.1 Latar Belakang**

Jantung adalah organ yang memompa darah ke seluruh jaringan tubuh. jika kerja pemompaan jantung menjadi tidak efisien, organ vital seperti otak dan ginjal menderita dan jika jantung berhenti bekerja, kematian terjadi dalam beberapa menit. Organisasi Kesehatan Dunia (WHO) memperkirakan 12 juta kematian terjadi di seluruh dunia, setiap tahun karena penyakit jantung. Oleh sebab itu , pada tahun 2012 Dr Bhuvaneswari membangun suatu Sistem prediksi penyakit kardiovaskular menggunakan algoritma genetika dan *neural network*. Yang bertujuan untuk memprediksi risiko dari penyakit kardiovaskular yang menyerang jantung dan pembuluh darah. Pada penelitian yang dilakukan Dr Bhuvaneswari, menggunakan dataset dari repositori UCI (*University of California, Irvine*) yang terdiri dari 303 kasus penyakit jantung(*sample*) dengan lima kelas yang terbentuk yaitu *absent , low , medium, high, serious*. Dengan begitu bisa memprediksi tingkat keparahan penyakit jantung ini, Kombinasi ini telah menghasilkan TP sebesar 94.55 % , FP sebesar 6.25 % , TN sebesar 93.75 %, FN sebesar 5.45 % dengan akurasi 94.17 % [1] .

Kemudian pada tahun 2013 Dhana *et al ,*membangun Sistem Prediksi Penyakit Jantung menggunakan metode *Naive Bayes*. Dengan menggunakan dataset yang sama yaitu dataset yang berasal dari UCI (University of California, Irvine) , pada penelitian tersebut kelas yang terbentuk ada lima juga *yaitu no, low, average, high dan jvery high*. Namun kali ini menggunakan pendekatan pengklasifikasi untuk deteksi penyakit jantung dan menunjukkan bagaimana *Naive Bayes* dapat digunakan untuk tujuan klasifikasi. Ada beberapa percobaan yang dilakukan pada penelitian tersebut menggunakan 303 sample , percobaan pertama menghasilkan akurasi sebesar 88.76 %, percobaan kedua menghasilkan akurasi sebesar 89.58 %, dan yang terakhir menghasilkan akurasi sebesar 88.96%. Pada setiap percobaan dilakukan perombakan data uji dan data latih, sehingga akurasi yang dihasilkan juga berbeda – beda. Dengan tiga percobaan diatas, akurasi yang dihasilkan tidak jauh berbeda, sehingga pengklasifikasian penyakit jantung menggunakan *Naïve bayes* cukup bagus [2]. Selanjutnya pada tahun yang sama yaitu 2013, Ratna *et al* juga membangun suatu system Prediksi Penyakit Jantung menggunakan algoritma genetic untuk pemilihan atribut *reduced optimal* yang bertujuan untuk memprediksi risiko penyakit jantung. Karena Organisasi Kesehatan Dunia pada tahun 2012 melaporkan bahwa 11,8 % dari total kematian global (di Amerika Serikat ) disebabkan untuk Penyakit Kardio Vaskular . Dengan menggunakan dataset yang sama yang bersumber dari UCI ( *University of California Irvin*), penelitian yang di lakukan oleh Ratna ialah suatu kombinasi antar algoritma genetic*, naïve bayes* dan *Decision Tree* . Namun penelitian tersebut algoritma genetic dan *native bayes* digunakan pada proses *reduced attribute set*. Sehingga terdapat dua hasil yang diporoleh dari *naïve bayes* dan *decision tree*, hasil menunjukkan bahwa *naïve bayes* lebih baik dari pada *decision tree* . Karena *native bayes* lebih konsisten dan waktu yang diberikan lebih cepat dari pada *decision tree* walaupun akurasi *naive bayes* lebih kecil dari pada *decision tree*. Akurasi yang dihasilkan oleh naïve bayes sebesar 96.5 % dengan waktu 0.02 detik, sedangkan *Decision tree* menghasilkan akurasi sebesar 99.2 % dengan waktu 0.09 detik [3].

Kemudian pada tahun yang sama juga Bakhtiar melakukan riset menggunakan Algoritma *Neural Network* ( NN) yang dimana neural network itu sendiri suatu usaha untuk meniru fungsi otak manusia. Dengan menggunakan dataset yang sama yang bersumber dari UCI namun data yang diperoleh berbeda dengan sebelumnya yaitu sebanyak 573 pasien, yang dimana 259 pasien terdeteksi sakit dan 314 pasien terdeteksi sehat. Penelitian tersebut bertujuan untuk memprediksi secara akurat adanya penyakit jantung karena jumlah kematian akibat penyakit jantung 959.227 pasien, yakni 41,4 % dari seluruh kematian atau setiap hari 2600 penduduk meninggal akibat penyakit jantung. Dari tiga belas attribute yang ada pada dataset dikurangi menjadi 6 attribut, menggunakan pencarian genetic. Selanjutnya dilakukan pengklasifikasian menggunakan tiga pengklasifikasi seperti *Naive Bayes*, Klasifikasi dengan *clustering* dan *decision tree* digunakan untuk memprediksi diagnosis pasien dengan akurasi yang sama sebagaimana diperoleh sebelum pengurangan jumlah atribut . Hasil dari pengujian model yang dilakukan adalah menentukan prediksi penyakit jantung dengan *neural network* untuk menentukan nilai *accuracy* dan AUC. Metode pengujiannya menggunakan *cross validation* dan menghasilkan akurasi sebesar 91.45 % , nilai presisinya sebesar 92.79 % , nilai recall sebesar 94.27 % dan untuk nilai AUC sebesar 0.937 . Dengan akurasi sebesar itu, bisa disimpulkan bahwa *neural network* memberikan pemecahan untuk permasalahan penyakit jantung lebih akurat [4].

Selanjutnya pada tahun berikutnya yaitu tahun 2014 Negar *et al ,*meneliti perbandingan yang di hasilkan menggunakan metode penggelompokan dengan menggunakan tambahan PCA (*Principal Component Analysis*) yang dimana penelitian tersebut bertujuan untuk mengetahui apakah PCA (Principal Component Analysis) memiliki pengaruh atau tidak terhadap metode pengelompokan. Metode yang dipakai ialah K-means dan FCM (*Fuzzy C - Means*) karena Metode fuzzy C-means dan K-means keduanya merupakan metode *unsupervised* *clustering*. Sedangkan untuk dataset yang digunakan adalah dataset dari repositori UCI (University of California, Irvine), berbeda dari sebelumnya data yang digunakan berjumlah 303 *sample* yang terdiri dari 164 sample orang sehat dan 139 untuk sample orang yang tidak sakit dengan menggunakan 13 attribut . Akurasi yang dihasilkan tanpa menggunakan PCA pada K-means sebesar 81 % dengan waktu 0.28 detik dan FCM tanpa menggunakan PCA sebesar 80 % dengan waktu 0.45 detik. Sedangkan akurasi yang di hasilkan menggunakan PCA pada K-means sebesar 87 % dengan waktu 0.1 detik, dan pada FCM dengan menggunakan PCA sebesar 82 % dengan waktu 0.32 detik. Hasil percobaan menunjukkan bahwa, algotima pengelompokan seperti K – mean dan FCM menggunakan tambahan PCA dapat meningkatkan kinerja metode pengelompokan, seperti meningkatkan akurasi dan mempercepat waktu komputasi [5].

Lalu pada tahun yang sama, Titi membuat suatu sistem cerdas yang dimana bertujuan untuk deteksi dini penyakit jantung menggunakan algoritma *decision tree* dan *fuzzy clustering*. Karena berdasarkan data yang dikeluarkan oleh PBB pada tahun 2005, penderita penyakit jantung dan pembuluh darah memilikiprosentase sebesar 29 % atau sebanyak 17,1 juta pasien meninggal dunia setiap tahunnya di seluruh dunia.Berdasarkan data dari Yayasan Jantung Indonesia, Indonesia sendiri memiliki prosentase penderita penyakit jantung sebesar 26,8 %. Jumlah kasus ini mengalami peningkatan setiap tahunnya dengan usia penderita yang semakin muda. Deteksi dini pada penyakit ini sebenarnya dapat dilakukan melalui diagnosa yang tepat dari seorang dokter. Dengan kombinasi algoritma itu, diterapkan pada dataset yang sama yaitu dataset dari UCI dan pembagian untuk proses uji coba yaitu berjumlah 30 data atau 10% dari jumlah keseluruhan data yang berjumlah sebanyak 297 sample. Algoritma decision tree digunakan untuk menentukan diagnosa penyakit jantung, sedangkan algoritma *fuzzy clustering* ( *Fuzzy C-Means* ) digunakan untuk klasterisasi stadium penyakit jantung yang diderita oleh pasien. Hasil yang didapat perbandingan nilai *true positif* dengan *false positif* sebesar 0.878 dalam bentuk nilai AUC. Nilai AUC tersebut sudah termasuk dalam kategori *good classification.* Sedangkan algoritma fuzzy C- means menggolongkan data ke dalam dua klaster dengan nilai SSE ( fungsi objektif) sebesar 235716,93567 . penggolongan ke dalam dua klaster dengan nilai SSE tersebut merupakan nilai fungsi objektif yang minimal dan tidak melebihi batas eror yang telah ditetapkan [6] .

Selanjutnya pada tahun 2015 Randa *et al*,menganalisis kumpulan data penyakit jantung arteri koroner. Tujuannya adalah untuk menerapkan integrasi hasil analisis pembelajaran mesin yang diterapkan pada kumpulan data berbeda yang menargetkan penyakit CAD (*Coronary artery disease*). Dengan dataset yang berbeda seperti dataset yang bersumber dari Cleveland heart disease , Hungarian heart disease , V.A heart disease , Statlog project dengan jumlah attribut yang sama yaitu 14 attribute namun jumlah data yang berbeda - beda. Randa *et al* menganalisis apakah ada perbedaan antara data yang dikumpulkan dengan yang terpisah. Metode yang di pakai untuk penelitian tersebut antara lainnya algoritma C4.5 dan Fast Decision Tree. Akurasi yang dihasilkan jika menggunakan dataset dari repositori UCI atau Clevelend pada algoritma C4.5 sebesar 78.54 % sedangkan akurasi yang dihasilkan menggunakan *Fast Decision Tree* sebesar 77.55 % . Kemudian akurasi dari dataset Hungarian pada algoritma C 4.5 sebesar 78.57 % , sedangkan akurasi menggunakan algoritma *fast decision tree* sebesar 78.23 % . Lalu akurasi dari dataset V.A mengguanakan algoritma C 4.5 sebesar 71.5 %, sedangkan jika menggunakan algoritma *fast decision tree* sebesar 69.5 %. Dan yang terakhir akurasi dari dataset Statlog jika menggunakan algoritma C 4.5 sebesar 76.6 %, sedangkan jika menggunakan algoritma *fast decision tree* sebesar 76.6 %. Hasil dari penelitian tersebut akurasi klasifikasi dari kumpulan data yang dikumpulkan adalah (77.50 %, 78.06 %) lebih tinggi dari rata-rata akurasi klasifikasi dari semua dataset terpisah yaitu (76.30 %, 75.48 %) [7] .

Kemudian pada tahun 2016 Agung *et al,* membangun suatu sistem prediksi menggunakan algoritma KNN(*K-Nearest Neighbor*) dengan pengurangan attribut. Dataset yang digunakan dari repositori UCI (*University of California, Irvine*), dari 13 attribute yang ada pada dataset hanya 8 attribute yang akan digunakan. Selain bertujuan untuk memprediksi penyakit jantung, pada penelitian tersebut juga meneliti apakah dengan pengurangan attribut akurasi yang dihasilkan akan lebih besar. Akurasi yang dihasilkan menggunakan algoritma KNN dengan pengurangan attribut sebesar 81.85 %, sedangkan jika tidak dikurangi menghasilkan akurasi sebesar 80.61 %. Jika dibandingkan dengan metode yang lainnya, akurasinya di bawah dari KNN. Seperti halnya Naïve bayes yang menghasilkan akurasi 74.49 % jika attributenya dikurangi dan 79.93 % jika tidak dikurangi. Kemudian ada metode Decision tree menggunakan algoritma *A Classification and* *Regression Tree*(*CART*), menghasilkan akurasi sebesar 80.27 % jika attributnya dikurangi dan 79.93 % jika tidak dikurangi. Sehingga didapat hasil bahwa KNN dengan pengurangan attribut dapat menghasilkan akurasi yang lebih baik [8] .

Pada tahun 2021 Shadab *et al,* mengusulkan arsitektur *deep learning*  menggunakan CNN (*Convolutional Neural Network*) untuk memprediksi penyakit jantung. Tujuannya ialah untuk diagnosis dini penyakit jantung dengan menggunakan sistem berbasis computer dengan CCN 1D. algoritma CNN ini mirip dengan algoritma ANN yang dimana kedua algoritma tersebut terdiri dari neuron yang mengoptimalkan diri yang dilatih untuk melakukan tugas tertentu. Hal itulah yang menyebabkan pengembangan CNN 1D yang dapat beroperasi pada dataset satu dimensi atau waktu data seri. Pada penelitian tersebut Dataset yang digunakan ialah Dataset penyakit jantung Cleaveland atau UCI, pada penelitian kali ini dibagi menjadi 2 yaitu 163 contoh negative dan 140 contoh positif dengan 13 attribut yang diinputkan untuk klasifikasi. Akurasi setelah setiap *epoch* menggunakan *ADAM optimizer* tanpa dropout sebesar 98.9 % pada set pelatihan dan 90.32 % pada set pengujian. Sedangkan akurasi setelah setiap *epoch* menggunakan *ADAM optimizer* sebesar 97.79 % pada set pelatihan dan 96.77 % pada set pengujian. Dengan akurasi sebesar itu maka CNN 1D dapat dikatakan cukup baik untuk memprediksi penyakit jantung baik dengan memakai optimizer maupun tidak memakai optimizer [9].

Tambahan lagi pada tahun yang sama, jefri *et al* meneliti untuk mendeteksi penyakit jantung. Pada penelitian tersebut menggunakan pendekatan *mechine learning*  dengan algoritma *logistic regression*, namun pada penelitian tersebut juga diterapkan pada metode lainnya seperti C.50 , SVM, dan Neural Network. Dimana algoritma *logistic regression* memakai fungsi logistik untuk menghasilkan *binary* atau nol dan satu sebagai penentuan klasifikasi. Dataset yang digunakan berjudul Heart Disease UCI dimana dataset tersebut didapat dari Cleaveland Heart Disease Database dan dipublikasikan pada Kaggle. Data yang digunakan untuk testing sebesar 20 % dari total data sehingga didapatkan sebesar 62 data untuk *testing*. Pada data *training*, metode logistik regresi mempunyai nilai *sensitivity* yang paling tinggi yaitu 88.54 % dibanding metode lainnya. Pada data *testing*, metode logistik regresi mempunyai nilai kekhususan yang paling tinggi yaitu sebesar 87.50 % dibanding metode lainnya [10].

Berdasarkan uraian diatas bahwa metode yang lebih baik dalam hal memprediksi penyakit jantung ialah metode *naïve bayes* kombinasi *algoritma genetic* dengan akurasi sebesar 96.5 % dan waktu komputasi 0.02 detik. Namun, CNN 1D memiliki akurasi yang lebih besar yaitu 97.79 % pada set pelatihan dan 96.77 % pada set pengujian dengan menggunakan *ADAM optimizer*. Kemudian akurasi yang didapatkan jika tidak menggunakan *ADAM optimizer* sebesar 98.9 % pada set pelatihan dan 90.32 % pada set pengujian.

Oleh karena itu, didapatkan suatu gambaran yang jelas untuk melakukan penelitian mengenai faktor-faktor apa saja yang mempengaruhi algoritma CNN (*Convolutional Neural Network*) dalam mengklasifikasikan penyakit jantung. Pada proposal ini akan menggunakan metode klasifikasi CNN dengan data yang bersumber dari UCI (*University of California*, *Irvine*), namun menggunakan optimizer yang berbeda - beda. Dengan usulan penelitian ini diharapkan sistem dapat mengklasifikasikan penyakit jantung yang akurat.

**1.2 Permasalahan**

Berdasarkan latar belakang di atas, maka permasalahan yang dihadapi adalah akurasi dalam mengklasifikasikan penyakit jantung . Pada proposal ini akan mengembangkan model untuk mengklasifikasikan penyakit jantung dari data yang ada.

**1.3 Batasan Masalah**

Berdasarkan latar belakang dan permasalahan diatas, maka batasan – batasan masalah pada penelitian ini sebagai berikut :

1. Optimizer yang dipilih yaitu AdaGrad, SGD dan Nadam.
2. Implementasi dan penggunaan optimizer.

**1.4 Tujuan**

Berdasarkan latar belakang dan permasalahan diatas, tujuan dari penelitian ini adalah untuk mengetahui faktor - faktor yang mempengaruhi akurasi pada algoritma CNN (*Convolutional Neural Network*) dalam mengklasifikasi penyakit jantung khususnya optimizer.

* 1. **Pertanyaan Penelitian**

1. Berapa akurasi klasifikasi penyakit jantung yang dihasilkan menggunakan model CNN dari setiap optimizer yang berbeda?
2. Manakah optimizer yang terbaik untuk model CNN dalam hal klasifikasi penyakit jantung?

**DAFTAR PUSTAKA**

[1] N. G. B. Amma, “Cardiovascular disease prediction system using genetic algorithm and neural network,” 2012. doi: 10.1109/ICCCA.2012.6179185.

[2] D. S. Medhekar, M. P. Bote, and S. D. Deshmukh, “Heart Disease Prediction System using Naive Bayes,” *Int J Enhanc Res Sci Technol Eng*, vol. 2, [Online]. Available: www.erpublications.com

[3] S. Ratnakar, K. Rajeswari, and R. Jacob, “PREDICTION OF HEART DISEASE USING GENETIC ALGORITHM FOR SELECTION OF OPTIMAL REDUCED SET OF ATTRIBUTES,” 2013.

[4] B. Rifai, “ALGORITMA NEURAL NETWORK UNTUK PREDIKSI PENYAKIT JANTUNG,” 2013. [Online]. Available: http://www.bsi.ac.id

[5] N. Ziasabounchi and I. N. Askerzade, “A Comparative Study of Heart Disease Prediction Based on Principal Component Analysis and Clustering Methods.” [Online]. Available: http://tjmcs.matder.org.tr/@TJMCShttp://www.matder.org.tr/

[6] T. M. Tatilah, “PEMBUATAN SISTEM CERDAS UNTUK DETEKSI DINI PENYAKIT JANTUNG DENGAN MENGGUNAKAN ALGORITMA DECISION TREE DAN FUZZY CLUSTERING.”

[7] R. El-Bialy, M. A. Salamay, O. H. Karam, and M. E. Khalifa, “Feature Analysis of Coronary Artery Heart Disease Data Sets,” in *Procedia Computer Science*, 2015, vol. 65, pp. 459–468. doi: 10.1016/j.procs.2015.09.132.

[8] I. Ketut, A. Enriko, M. Suryanegara, and D. Gunawan, “Heart Disease Prediction System using k-Nearest Neighbor Algorithm with Simplified Patient’s Health Parameters”.

[9] S. Hussain, S. Barigidad, S. Akhtar, and M. Suaib, “Novel Deep Learning Architecture for Heart Disease Prediction using Convolutional Neural Network.”

[10] J. Junifer Pangaribuan and H. Tanjaya, “MENDETEKSI PENYAKIT JANTUNG MENGGUNAKAN MACHINE LEARNING DENGAN ALGORITMA LOGISTIC REGRESSION,” 2021.